

Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay

A. Jarvis^{1,2,*}, K. Williams³, D. Williams¹, L. Guarino¹, P.J. Caballero⁴ and G. Mottram²

¹International Plant Genetic Resources Institute (IPGRI), Cali, Colombia; ²Centro Internacional de Agricultura Tropical (CIAT), Cali, AA6713, Colombia; ³United States Department of Agriculture (USDA), Agricultural Research Service, Beltsville, MD, USA; ⁴Dirección de Investigación Agrícola, Ministerio de Agricultura, Paraguay; *Author for correspondence (e-mail: a.jarvis@cgiar.org)

Received 24 March 2003; accepted in revised form 3 October 2003

Key words: *Capsicum*, Conservation, *Ex situ*, Geographic information systems, Habitat loss, *In situ*, Paraguay, Plant genetic resources, Principal components analysis, Rare species, Species distribution, Wild pepper

Abstract

This paper presents an effective method for prioritizing areas within a country for acquisition of germplasm of a crop gene pool for *ex situ* conservation. The method was applied to the rare wild pepper species, *Capsicum flexuosum* Sendtn., in south-east Paraguay. A model to prioritize areas for collecting germplasm was constructed by combining (1) a prediction of the species' geographic distribution based on the climate at previous collection points, (2) the distribution of forest margins (the species' natural habitat) and (3) areas accessible by road. The model was then tested in the field by visiting 20 sites having both high and low predicted probability of occurrence of *C. flexuosum*. Six new populations were found, representing a significant improvement over two previous collecting missions for the species in the same region, undertaken without the use of GIS targeting. Using the most optimistic analysis of model performance, *C. flexuosum* was found at five out of seven points predicted to harbour the species and not found at four of five points predicted not to harbour the species. The model was then improved by the use of higher resolution climate surfaces. It is recommended that future explorers use more recent and higher resolution satellite images to locate suitable habitats. The method is replicable for different species in different geographic regions and is offered as a means of optimizing efficiency in financially constrained, national plant genetic resources programs.

Abbreviations: GRIN – Germplasm Resources Information Network; IGBP – International Geosphere-Biosphere Program; NPGS – National Plant Germplasm System

Introduction

The well-documented global loss of biodiversity impacts plant genetic resources as severely as other species. The Report on the State of the World's Plant Genetic Resources (FAO 1996), which summarizes reports submitted by 151 coun-

tries, states that “recent losses of diversity have been large, and that the process of ‘erosion’ continues.” The preservation of shrinking crop gene pools, critical resources for agricultural sustainability, is of the utmost importance. Many of the wild species comprising crop gene pools have barely been observed, collected or described (Guarino et al.

2001). These species are precisely those which are most threatened because of their rarity, endemism and a general lack of knowledge about them.

The gene pool of *Capsicum* is comprised of about 30 species, all of which are endemic to the Americas. Five species of *Capsicum* were independently domesticated in prehistoric times for their edible, pungent fruits. Commonly known as red pepper, "chile" or "aji," *Capsicum* fruits are now the most widely consumed spice and vegetable crop in the world. Traditional cultivars of *Capsicum* and their wild relatives are found throughout the Americas, from the continental United States to Argentina. One of the lesser-known species in the *Capsicum* gene pool is *Capsicum flexuosum* Sendtn., a wild species native to south-eastern Paraguay, south-western Brazil and north-eastern Argentina. In 1950, Armando Hunziker, the internationally recognized specialist in *Capsicum* systematics, mistakenly confused *C. flexuosum* in Paraguay with *C. schottianum*, a species endemic to Brazil (Hunziker 1950). In a later publication, after conducting fieldwork in coastal Brazil and studying the holotype of *C. schottianum*, he corrected his error of 1950 and clarified the distributions of the two species (Hunziker 1998).

The first known germplasm accession of *C. flexuosum* was collected during a mission targeting several *Capsicum* species in Paraguay in 1998; however, less than 20 seeds were recovered. Despite 11 days of extensive searching that covered much of south-eastern Paraguay, the population was found in a location where a herbarium specimen was collected 18 years previously. Other exact locations where herbarium specimens of *C. flexuosum* had previously been collected were visited but, in most cases, the natural habitat (forest margins) had been severely altered by deforestation, overgrazing or intensive agriculture, and the species could no longer be found. To obtain additional germplasm for *ex situ* conservation of this rare and evidently threatened crop relative, a second exploration for *C. flexuosum* was undertaken in 2001, with much the same result. One population was found in the Itabó Biological Reserve in close proximity to previously recorded locations of herbarium specimens. Only 27 viable seeds were recovered. The explorers also observed that the habitat destruction noted previously was continuing unabated in most of the unprotected areas explored.

Attempts to increase the first germplasm samples of *C. flexuosum* under greenhouse conditions suffered from poor seed set (Paul Bosland, personal communication 2002), likely caused by self incompatibility, suggesting the need for larger samples of germplasm. Some of the projected uses of this germplasm are experimental crosses and biochemical and molecular studies to elucidate the relationships between *C. flexuosum* and other wild and cultivated *Capsicum* species, and to help determine the value of the species as a genetic resource for crop improvement.

With so little known about the species, identifying the precise geographic range of *C. flexuosum* is difficult but vital if more germplasm is to be collected and conserved. Plant collectors often use vegetation maps to locate suitable regions for collection, but this is both subjective and reliant on the availability and quality of these maps. The geographic range can be modelled based on the sites of existing collections, or through the manipulation of associated variables that affect the geographic distribution. Anderson et al. (2002) state that shaded outline maps ranging between and beyond known localities are likely to overestimate species distribution, while dot maps of known localities portray species distribution conservatively. Geographic bias in collecting efforts centring along roads (Hijmans et al. 2000) creates further error in approximating species range using these methods. Species distribution modelling of associated variables presents a means of extrapolating species range from point localities to a wider region, while minimizing the risk of over- or under-estimation (Franklin 1995). Guisan and Zimmerman (2000) discuss some of the applications of species distribution modelling and the various modelling algorithms that have been applied to the problem. Many of these methods use climatic variables as the principal drivers of geographic distribution (Walker and Cocks 1991; Franklin 1995; Guisan and Zimmerman 2000). Jones et al. (1997) used the FloraMap computer program to predict the geographic distribution of wild bean (*Phaseolus vulgaris* L.) based on the distribution of germplasm and herbarium specimens. The results correctly predicted areas where wild bean had not been collected but was reported to occur in the literature. Jarvis et al. (2002, 2003) used the same method to assess the conservation

status of wild peanuts in South America. Segura et al. (2002) also used FloraMap to map the geographic distribution of five species of *Passiflora* and successfully guided germplasm collecting in Ecuador. Guarino et al. (2001) provide a general discussion on the application of species distribution models in the conservation and use of plant genetic resources.

Here, we report the results of the third consecutive exploration for *C. flexuosum* and an assessment of the GIS-based methodology used in its planning and execution. It is hoped that the methodology will present germplasm collectors with a better means of targeting their conservation efforts, to improve the success of their endeavours, saving time and money in the process.

Materials and methods

Geographic information systems (GIS) tools were used to analyse information on the known locations of past and present populations of *C. flexuosum*, together with other geo-referenced datasets. The model combined:

- a prediction of the potential distribution of *C. flexuosum* based on climatic adaptations determined from previous collecting sites;
- a map of the distribution of forest margins, the known preferred habitat of the species; and
- a map of 4-km wide buffers along roads to restrict collecting localities to within walking distance of roads.

A database of herbarium specimens and germplasm accessions of *C. flexuosum* was collated. Information from herbarium specimens of *C. flexuosum* in three herbaria in Paraguay (PY, FCQ, AS), five in Argentina (BA, BAB, CTES, LIL, SI), and one in the United States (MO), was combined with information on germplasm accessions from the US National Plant Germplasm System into a database that includes geographic coordinates for each observation. The database contains 19 unique populations of the species, 13 of which originate from Paraguay and six from Argentina.

Species distribution modelling

FloraMap (Jones and Gladkov 1999) was used to develop climatic models for predicting the

distribution of *C. flexuosum* in Paraguay. FloraMap predicts the distribution of organisms in the wild when little is known of the physiology of the species involved. It is assumed that the climate at the points of observation and/or collection of a species is representative of the environmental range of the organism. The climate at these points is used as a calibration set to compute a climate probability model.

FloraMap uses climatic data from a 10-minute grid (which corresponds to 18×18 km at the equator) derived from observations from over 10 000 meteorological stations in Latin America, 130 of which are located in Paraguay. A simple interpolation algorithm, based on the inverse square of the distance between stations and the interpolated point, is used. For each interpolated pixel, the five nearest stations are used in the inverse distance equation. The climatic variables included are the monthly averages for temperature, rainfall and diurnal temperature range. Mean temperature is standardized with elevation using the NOAA TGP-006 (NOAA 1984) digital elevation model and a lapse rate model (Jones 1991). Rainfall and diurnal temperature range were found to be independent of elevation. A 12-point Fourier transform is applied to each variable to adjust for geographic differences in the timing of major seasons. For further information, see Jones et al. (1997, 2002).

For each accession, the 36 climate variables (12 monthly means for temperature, rainfall and diurnal temperature range) were extracted for the pixel in which the accession is located, and a principal components analysis (PCA) applied to identify a smaller number of variables that account for the bulk of the variance in climates among the accession locations. The PCA is performed on the variance-covariance matrix since the Fourier analysis has transformed the variables to comparable scales. A multivariate-Normal distribution is fitted to the principal component scores so that a probability of belonging to the distribution can then be calculated for all pixels on a continent. The result is a probability surface for the study area in south-eastern Paraguay (Figure 1). It should be noted that this merely maps the potential climatic envelope where an organism could exist and does not account for factors such as edaphic preferences, dispersal mechanisms, anthropogenic impacts or ecological habitat preferences. With so little

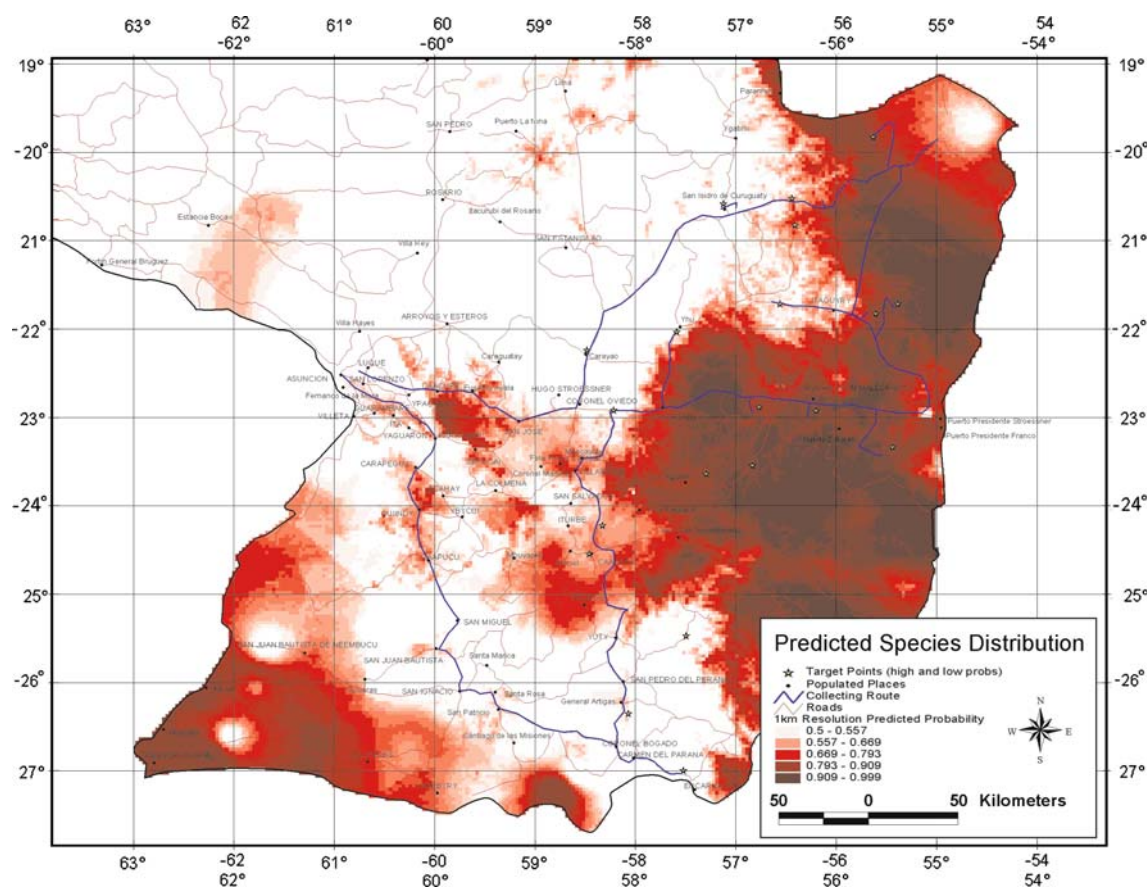


Figure 1. Predicted species distribution of *C. flexuosum* Sendtn. in Paraguay using the 1-km climate grid in FloraMap.

known about the species, it is impossible to assess to what extent other factors may limit this distribution.

The original analysis of species distribution of *C. flexuosum* was made on the 10-min grid in FloraMap, the results of which were taken into the field. After the collecting mission, the climate data were improved to a 1-min grid (about 2-km grid cell resolution) and the analysis repeated. The same meteorological station data were used to generate the new climate surfaces but were improved through regional modelling of the relationship between elevation and climate variables. The higher resolution grids capture more spatial variability in climate, especially in the more topographically complex south-eastern region of Paraguay. These climate data were then used to repeat the prediction of species distribution for *C. flexuosum*, using only the accession and herbarium records used in the original analysis.

Additional GIS layers

Two additional GIS layers were used to limit the search area, given the limited time and resources to extensively search all regions of the country. These layers were not, however, used in the statistical analyses. To identify areas of potentially suitable habitat, the International Geosphere–Biosphere Program (IGBP) Land Cover Classification (Belward 1996) 1-km land-cover data set (derived from advanced very high resolution radiometer satellite images) was used to locate forests in Paraguay. This data set is freely available from the United States Geological Survey (USGS) Global Land Cover Characteristics project (<http://edc-daac.usgs.gov/glcc/glcc.html>). The satellite images used to produce the data set were acquired for the 12-month period between April 1992–March 1993. In all, five forest land-covers (Deciduous Broadleaf Forest, Deciduous Needleleaf Forest, Evergreen

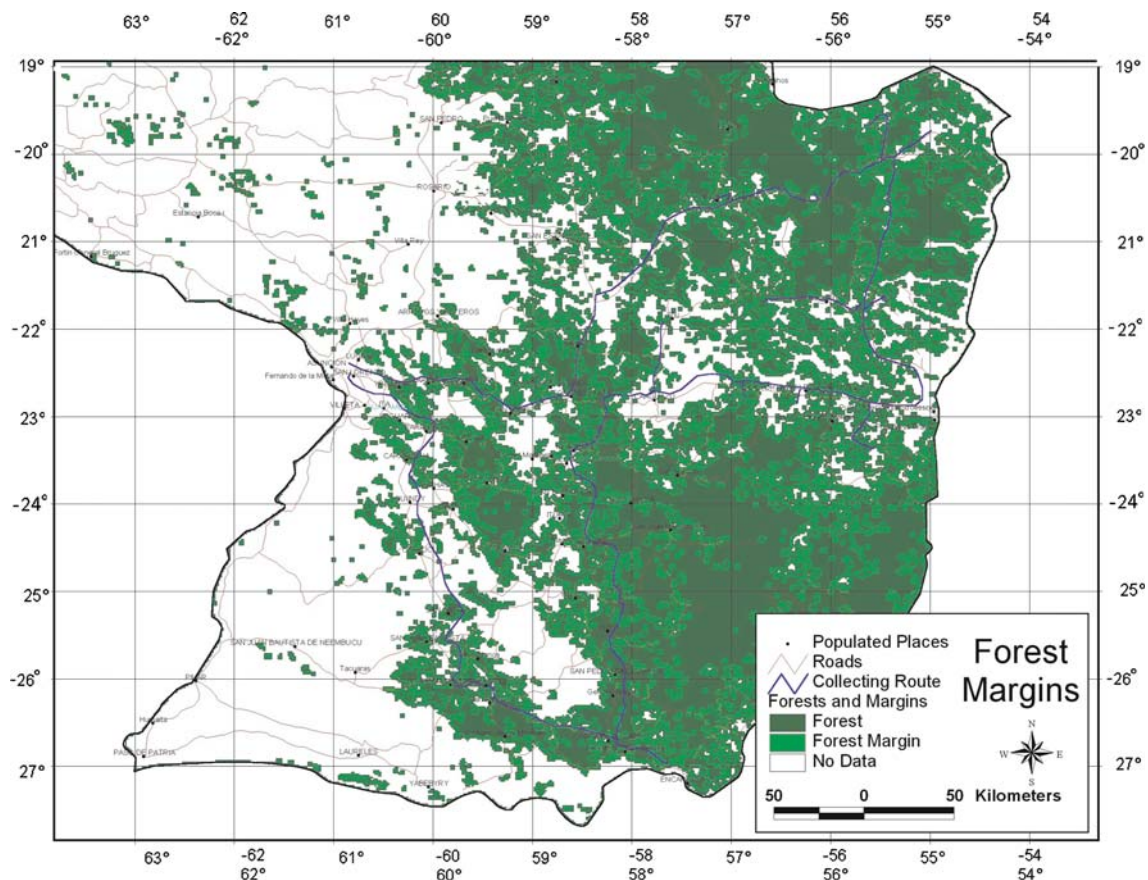


Figure 2. Distribution of forests and of forest margins in southern Paraguay, defined through use of the International Geosphere-Biosphere Program (IGBP) Land Cover Classification (Belward 1996).

Broadleaf Forest, Evergreen Needleleaf Forest and Mixed Forest) were joined to produce a single coverage of forests in Paraguay. Areas within 1 km of the forest edge were selected using ArcView 3.2 (Environmental Systems Research Institute [ESRI], Redlands, CA, USA) (Figure 2).

Finally, the Digital Chart of the World (ESRI 1992; freely downloadable GIS coverage available from the Penn State University Library Digital Chart of the World Server at <http://ortelius.maproom.psu.edu/dcw/>) was used to locate roads in the collecting region and a 4-km buffer around these roads was calculated in ESRI's ArcView 3.2 to locate accessible areas. The buffer zone was then applied to limit the extent of the collecting priority model to include only areas where collection was deemed feasible (Figure 3).

Field testing

An experiment was devised to validate and determine the practical usefulness of the FloraMap model and to ensure the objectivity of the test. Ten accessible points were randomly selected from areas where the model predicted the highest probability of encountering the target species and 10 additional accessible points were randomly selected from nearby areas with low predicted probability of supporting populations of *C. flexuosum*. All 20 points were displayed, undifferentiated, on a map of the collecting area and the precise coordinates (latitude and longitude) were provided to the plant explorers so that they could use a global positioning system (GPS) receiver to reach the precise location of each point. The explorers were not

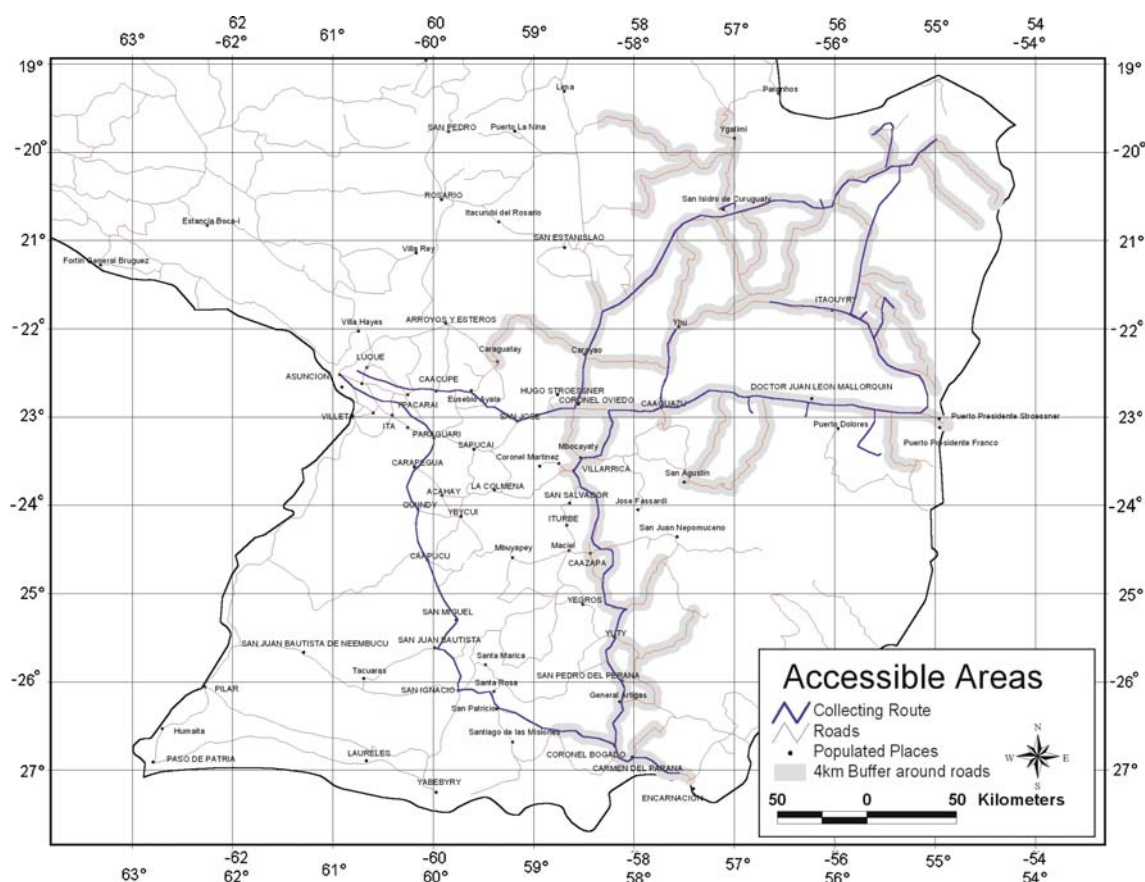


Figure 3. Areas defined as being accessible to plant collectors within the target region for *C. flexuosum*. The buffer around the major roads was 4 km.

informed which points represented areas of high or low predicted probability for the target species. Their task was to visit and explore each of the 20 points for *C. flexuosum* for a recorded period of time, describe the location, collect herbarium and germplasm specimens whenever possible and report their findings. At each point where the collectors stopped, the habitat was assessed as to whether it was likely to harbour populations of *C. flexuosum*. Additional information was recorded as to the approximate area searched and the number of collectors involved in the search. Detailed notes were taken as to the habitat and surrounding vegetation at each search site.

Evaluations of species distribution models typically use presence/absence data to test how well the prediction fits with reality (Fielding and Bell 1997). Confirming absence of a species at a given site presents a number of difficulties, especially if the

area is larger than can be feasibly sampled. This creates a degree of uncertainty in model evaluation for points of absence (Figure 4). Manel et al. (2001) conclude that Cohen's Kappa provides the most appropriate statistical evaluation of presence/absence validation. This form of evaluation is problematic in germplasm collections for which only species presence data are normally available but applicable in the case of this study because of the predetermined sampling strategy.

Validation of the models involved classifying species presence/absence, USGS land-cover, field-observed land-cover and the species probability distribution into 2×2 contingency tables for calculation of Cohen's Kappa statistic. Probabilities of >0.5 from FloraMap were assumed to indicate predicted presence of the species, whilst presence of forest margins in the USGS landcover data was assumed to indicate suitable habitat. To account

		Observed	
Predicted		Absence	Presence
	Absence	CORRECT MODEL RESULT UNCERTAIN True negative (or false negative?)	MODEL ERROR False negative
	Presence	MODEL ERROR UNCERTAIN False positive (or true positive?)	CORRECT MODEL RESULT True positive

Figure 4. Matrix of potential outcomes in the validation of a presence/absence model.

for instances of uncertainty in confirming the absence of the species (e.g., in the case of the 18-km grid cell predicted species distribution, the collectors can only feasibly search a small percentage of the total area of the grid cell), the statistical analysis was made for three scenarios in order to assess the range of potential model performances. The following three scenarios were considered, ranging from pessimistic (model performance potentially lowest) to optimistic (assumptions and ambiguities in recorded absences questioned in order to represent model performance more favourably):

- *Pessimistic scenario.* All recorded absences are assumed to be true.
- *Realistic scenario.* Recorded absences where the collectors observed no suitable habitat due to anthropogenic impacts were omitted from the analysis. All other observed absence sites where the habitat was deemed suitable were included as confirmed absences.
- *Optimistic scenario.* Same as realistic scenario, with additional filtering of sites depending on the uncertainty of the absence results. This was calculated by analysing data for the area searched and habitat characteristics of each site where *C. flexuosum* was not found. The following equation was used to estimate the degree of certainty in the absence of the species:

$$\text{Certainty(\%)} = (\% \text{ grid cell searched} / \% \text{ grid cell with suitable habitat}) * 100 \quad (1)$$

This equation produces the percentage of the total area searched that is considered suitable for

C. flexuosum. It fails to take into account collector error in not spotting the plant or occasions when the plant is under stress and has few or no leaves, making it impossible to notice, let alone identify. If the certainty was greater than 50%, the site was assumed to be a confirmed absence and if less than 50%, the site was dropped from the analysis.

Results and discussion

The result of the modelling highlights an area of 6460 km² in south-eastern Paraguay where *C. flexuosum* is predicted to occur (Figure 5).

During 8 days in the field from 15–22 March 2002, 20 sites were visited and explored for *C. flexuosum*, including 17 of the 20 predetermined points and three other unplanned sites (Figure 6, Table 1). Three predetermined points were inaccessible, either because of poor roads or the unmanageable distance of the target sites from the road. Populations of *C. flexuosum* were discovered at six of the predetermined locations, resulting in six herbarium and three germplasm collections, totalling over 160 seeds. All these populations were previously unknown and distant from existing germplasm or herbarium collection sites. It is clear that the GIS predictive model successfully guided the germplasm explorers to new populations of the target species and significantly increased the effectiveness of the exploration (compared with the limited success of previous collecting missions). New populations were discovered and more germplasm of *C. flexuosum* was obtained than on any previous exploration for this species.

The low number of data points compromises all statistical analyses but these provide some useful comparative measures of model performance. The comparison of habitat from the remote sensing data with the field collectors' perception of suitable habitats produced a Kappa statistic of 0.19, indicating a poor match. It is likely that the 1-km grid cell resolution and age of satellite imagery in the USGS land cover data set caused this discrepancy. South-eastern Paraguay has experienced high rates of deforestation in recent years, with the result that forest habitats identified in 1992–93 may now be severely degraded or lost.

Model performance steadily improves from the pessimistic to optimistic scenario for both the 18

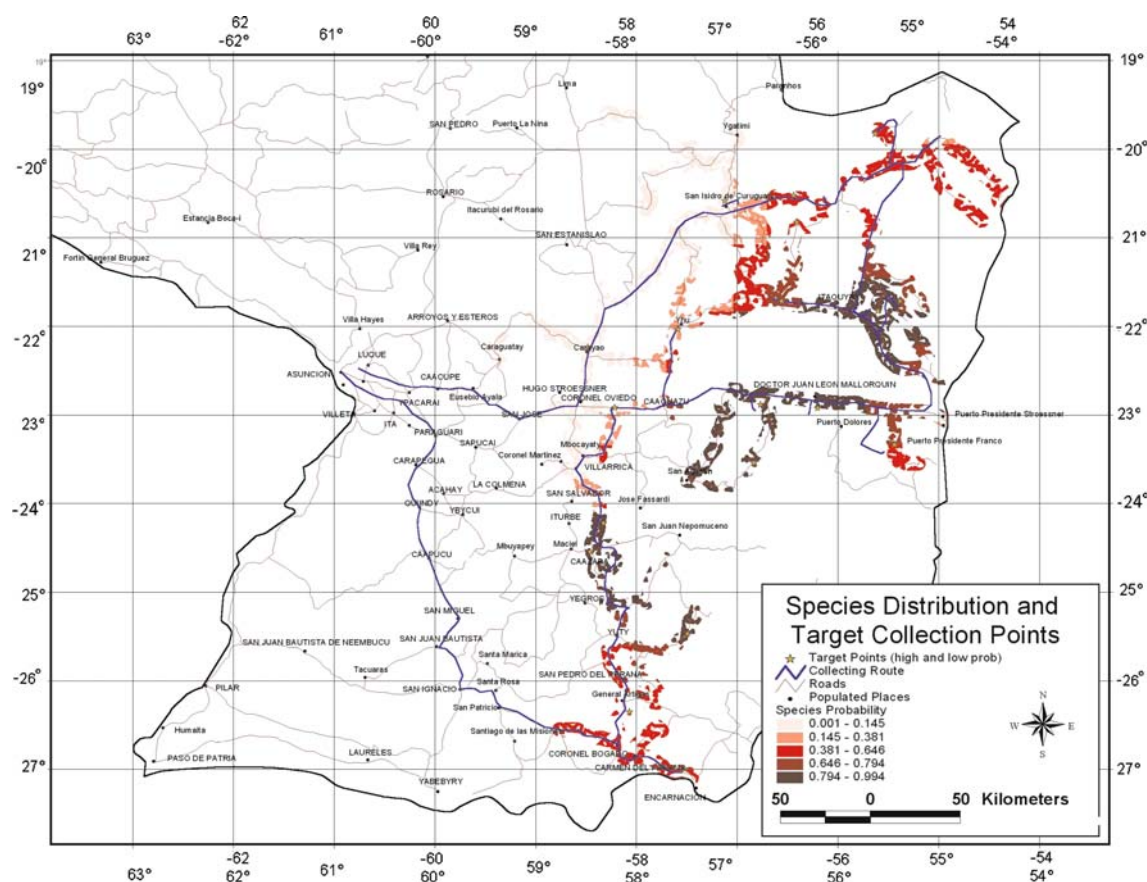


Figure 5. Priority areas for collecting in south-eastern Paraguay. The model combines the distribution of the species predicted by FloraMap with its potential habitat in forest margins and masks out areas not considered accessible by road or walking (4-km buffer).

and 1-km model (Tables 2 and 3). Using the realistic scenario, the 1-km model correctly predicted the presence/absence in 64% of sites, correctly predicting the absence in four out of five sites. The 1-km model performance is clearly better, with greater accuracy at predicting true negatives.

Cohen's Kappa was developed for medical applications, where Kappa values of 0.0–0.4 indicate slight to fair model performance, values of 0.4–0.6 moderate, 0.6–0.8 substantial and 0.8–1.0 almost perfect (Landis and Koch 1977; Manel et al. 2001). For all scenarios, the Kappa statistics had greater significance for the 1-km species distribution model than for the 18-km resolution model. This indicates that the higher grid cell resolution causes significant improvement in model success.

As expected, Kappa values steadily improve from the pessimistic to optimistic scenarios. They show low performance levels for all 18-km model

scenarios. However, for the 1-km model, the values show fair performance levels for the pessimistic and realistic scenarios and moderate levels of performance for the optimistic scenario.

Although subjective, the three scenarios provide the confidence limits to expect with this method of species distribution modelling. The pessimistic scenario is likely to underestimate the performance of the model. Some six sites where *C. flexuosum* was not found contained heavily altered habitats (often soybean fields), all of which were predicted to harbour the species. These sites prove nothing about the success of the species distribution model, which is based entirely on climate, but indicate the need for improved land-cover data to locate suitable natural habitats. The optimistic scenario dropped an additional seven sites from the 18-km model and two from the 1-km model, in which the absence of the species was deemed an uncertain result. In the

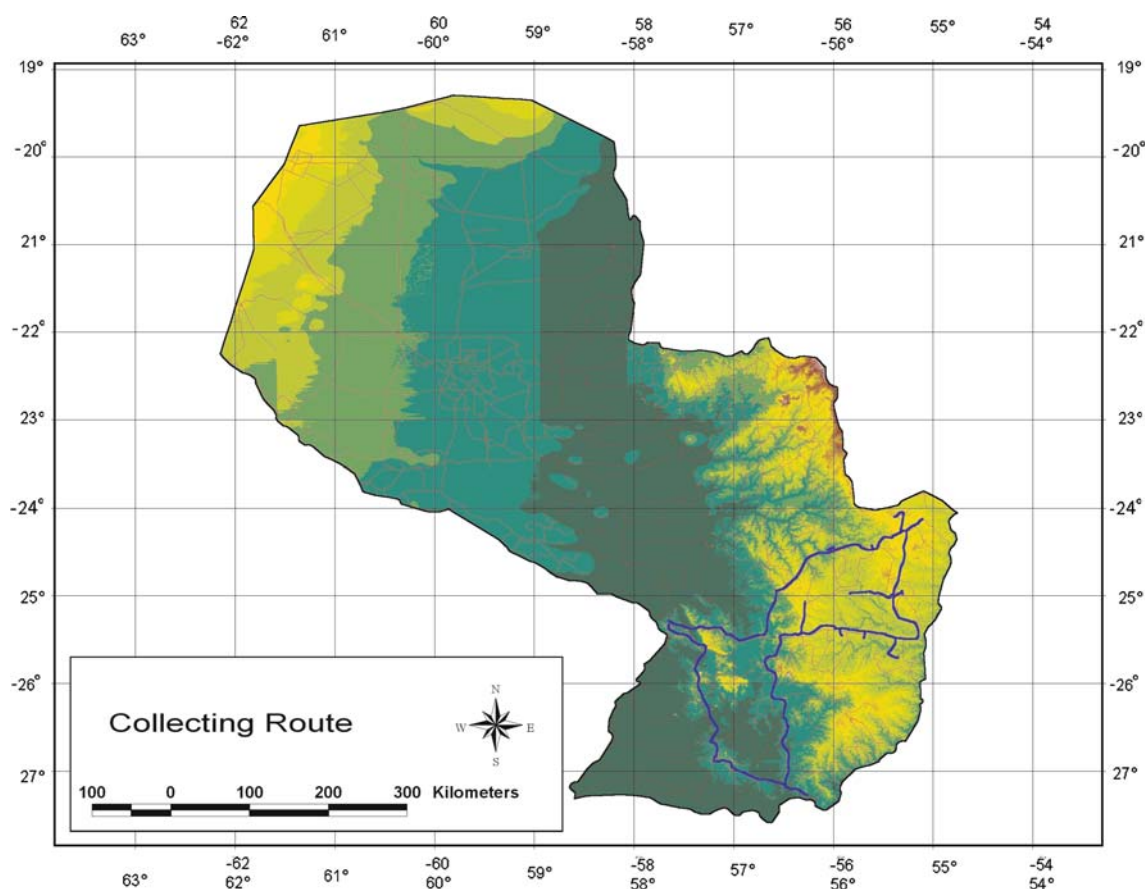


Figure 6. The final collecting route covered in south-eastern Paraguay in March 2002.

case of the 18-km model, confirming species absence was difficult because of the large area within which to search. For this reason, many points were dropped from the analysis. The 1-km grid cell resolution is a more feasible area within which to search and thus more absences were confirmed.

The true model performance is likely to be between the realistic and optimistic scenarios, offering around 60–65% accuracy for the 18-km model and 67–72% accuracy for the 1-km model. It should be noted that, compared to collections made without the aid of GIS targeting, this represents a significant improvement. In two similar collecting missions made prior to this study, only two populations were found and both were found in locations where herbarium samples had been taken previously. Here, six new populations were found in areas where *C. flexuosum* had not been previously found.

Conclusion

The fieldwork reported here was a successful validation of a GIS application designed to enhance the effectiveness of plant explorations. Using the high-resolution model and the optimistic scenario, *C. flexuosum* was found at five out of seven points predicted to harbour the species and not found at four of five points predicted not to harbour the species. This represents a considerable gain in efficiency compared to previous explorations during which the only germplasm collected was from the sites of previous herbarium specimens.

Validation of these methods requires more data than are commonly taken on germplasm collecting missions. In particular, collectors must make note of all sites that were searched for germplasm, including those where none was found. In all cases, variables such as the area searched, the time expended and the intensity of search (number of

Table 1. Characteristics of each location where collectors searched for *C. flexuosum* in south-eastern Paraguay.

Species presence	Search area (ha)	Time searching (min)	Altitude (m)	FloraMap 18-km grid probability	FloraMap 1-km grid probability	USGS ^a habitat of collection point	Collector perception of habitat suitability	Distance from road (km)	Longitude	Latitude
Yes	15	75	255	0.66	0.95	Savannah	Suitable	4.96	−55.26	−25.49
Yes	5	60	230	0.38	0.95	Evergreen Broadleaf Forest	Suitable	2.05	−54.88	−25.68
Yes	6	60	265	0.83	0.95	Savannah	Suitable	0.76	−54.85	−24.94
Yes	50	120	245	0.88	0.53	Evergreen Broadleaf Forest	Suitable	2.21	−55.44	−24.96
Yes	24	60	300	0.53	0.91	Evergreen Broadleaf Forest	Suitable	7.94	−55.01	−24.13
Yes	8	60	150	0.75	0.30	Evergreen Broadleaf Forest	Suitable	1.20	−55.71	−24.46
No	10	60	114	0.31	0.20	Evergreen Broadleaf Forest	Suitable	4.16	−56.02	−27.22
No	3	30	116	0.49	0.52	Evergreen Broadleaf Forest	Suitable	0.37	−56.21	−26.99
No	20	75	260	0.63	0.37	Savannah	Suitable	0.13	−55.91	−26.63
No	0	0	125	0.89	0.63	Evergreen Broadleaf Forest	Unsuitable	0.13	−56.40	−26.19
No	0	0	140	0.95	0.61	Evergreen Broadleaf Forest	Unsuitable	0.54	−56.33	−26.05
No	8	60	125	0.63	0.51	Savannah	Suitable	0.27	−56.5	−26.03
No	24	90	250	0.60	0.38	Evergreen Broadleaf Forest	Suitable	0.18	−56.27	−25.48
No	0	0	320	0.94	0.49	Savannah	Suitable	0.00	−55.95	−25.08
No	24	90	330	0.94	0.49	Savannah	Unsuitable	1.16	−55.97	−25.08
No	50	75	260	0.77	0.91	Evergreen Broadleaf Forest	Suitable	4.11	−55.45	−25.50
No	5	45	360	0.83	0.93	Evergreen Broadleaf Forest	Unsuitable	2.41	−54.95	−24.99
No	0.3	30	330	0.83	0.93	Evergreen Broadleaf Forest	Suitable	0.20	−54.95	−24.96
No	0	0	250	0.86	0.67	Evergreen Broadleaf Forest	Unsuitable	0.18	−55.37	−24.43
No	40	60	130	0.35	0.07	Evergreen Broadleaf Forest	Unsuitable	0.13	−56.39	−25.19

^aUSGS, United States Geological Survey.Table 2. Modelled presence/absence of *C. flexuosum* compared with field observations for the three scenarios.

Scenario	True positives		True negatives		False positives		False negatives		Performance % (% correct results)	
	18 km	1 km	18 km	1 km	18 km	1 km	18 km	1 km	18 km	1 km
Pessimistic (<i>n</i> = 20)	5	5	3	7	11	7	1	1	40	60
Realistic (<i>n</i> = 14)	5	5	2	4	5	4	1	1	54	64
Optimistic (18 km <i>n</i> = 7, 1 km <i>n</i> = 12)	5	5	0	4	1	2	1	1	71	75

Table 3. Kappa statistic for the comparison of modelled presence/absence with field observations for the three scenarios.

Scenario	Climate grid	
	18 km	1 km
Pessimistic ($n = 20$)	0.05	0.30
Realistic ($n = 14$)	0.09	0.34
Optimistic (18 km $n = 7$, 1 km $n = 12$)	0.05	0.50

people) at each site must be measured or approximated. In addition, observations of habitat to assess perceived suitability for the target species and percentage of the area covered by this habitat should be made. This detailed information allows the evaluation of the models using both presence and absence data and permits the approximation of degrees of uncertainty in absences (i.e., the probability that an observed absence may indeed be a false negative).

Despite these difficulties, the model was statistically shown to perform well. The species distribution model was significantly improved with a higher resolution climate database at 1 km. Based on our experience in the field, we make the following recommendations for improving the effectiveness and user-friendliness of the model:

- Obtain and verify the most recent road coverages available, perhaps through the use of recent LANDSAT TM satellite images.
- Obtain the most recent land-cover images and if possible use higher resolution images, such as those produced by LANDSAT TM (30-m resolution) to improve the targeting of habitats (cheaply available for areas of the tropics from the Tropical Rain Forest Information Center, <http://www.bsrsi.msu.edu/trfic/> or the MrSID Image Server, <http://zulu.ssc.nasa.gov/mrsid/mrsid.pl>).
- Reduce the buffer zone around roads from 4 to 2 km. When time is limited, 4 km proved to be an unrealistic distance to walk.
- Employ the improved FloraMap climate grids that have a resolution of 1 km.

While the climate envelope is likely useful for most wild species; the usefulness of vegetation will vary. Limiting the distance from roads is practical but may preclude finding the targets in some cases. Additional information, such as soils, hydrology or whatever else is deemed a relevant predictor, can be added where available.

Spatial prioritization methods, such as the example presented here, provide a tool to save time and money during collecting trips. The data needed are mostly freely available and the methods are replicable and cost-effective, given a computer and basic GIS software and skills.

Observations made at the 20 sites visited on the present exploration, together with those of previous collecting trips and the information gleaned from herbaria and gene banks, enable us to make some informed statements regarding the present conservation status of *C. flexuosum*, both *ex situ* and *in situ*. At the present time, germplasm samples from five natural populations have been deposited in the US National Plant Germplasm System (NPGS) for long-term *ex situ* conservation on behalf of the Paraguayan national authorities. Of these, one accession (WWQC 146) has been increased and assigned a Plant Introduction number (PI 631154) in the NPGS Germplasm Resources Information Network – GRIN (<http://www.ars-grin.gov/npgs/searchgrin.html>). The *in situ* conservation status of *C. flexuosum* in Paraguay continues to be threatened by habitat destruction.

Acknowledgements

This research and plant collecting was made possible by the National Geographic Society (Research and Exploration Grant No. 6434–99), the USDA-Agricultural Research Service, the International Plant Genetic Resources Institute (IPGRI), the Ministerio de Agricultura y Ganadería (Paraguay) and the Secretaría del Ambiente (Paraguay). Permits for collecting germplasm and herbarium specimens were obtained from the Secretaría del Ambiente in Paraguay. The authors would like to thank Victor Santander and Marizza Quintana for help during the collecting mission, German Lema for his aid in statistical analyses and Peter Jones for providing FloraMap and the 1-minute climate grid for Paraguay.

References

- Anderson R., Gomez-Laverde M. and Peterson A. 2002. Geographical distributions of spiny pocket mice in South

- America: Insights from predictive models. *Global Ecol. Biogeogr.* 11: 131–141.
- Belward A.S. (ed). 1996. The IGBP-DIS Global 1 km Land Cover Data Set (DISCover)-Proposal and Implementation Plans. IGBP-DIS Working Paper No. 13, Toulouse, France.
- ESRI (Environmental Systems Research Institute), 1992. Digital Chart of the World. ESRI, Redlands, CA, USA.
- Fielding A. and Bell J. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24: 38–49.
- Franklin J. 1995. Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geogr.* 19: 474–499.
- FAO, 1996. Report on the State of the World's Plant Genetic Resources for Food and Agriculture. FAO, Rome.
- Guarino L., Jarvis A., Hijmans R.J. and Maxted N. 2001. Geographic Information Systems (GIS) and the conservation and use of plant genetic resources. In: Engels J. (ed). *Managing Plant Genetic Diversity*, CAB International, Wallingford, UK, pp. 387–404.
- Guisan A. and Zimmerman N. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135: 147–186.
- Hijmans R., Garrett K., Huaman Z., Zhang D., Schreuder M. and Bonierbale M. 2000. Assessing the geographic representativeness of gene bank collections: The case of Bolivian wild potatoes. *Conserv. Biol.* 14(6): 1755–1765.
- Hunziker A.T. 1950. Estudios sobre Solanaceae. I. Sinopsis de las especies silvestres de *Capsicum* de Argentina y Paraguay. *Darwiniana* 9: 225–247.
- Hunziker A.T. 1998. Estudios sobre Solanaceae. XLVI. Los ajíes silvestres de Argentina (*Capsicum*). *Darwiniana* 36: 201–203.
- Jarvis A., Guarino L., Williams D., Williams K. and Hyman G. 2002. The use of GIS in the spatial analysis of wild peanut distributions and the implications for plant genetic resource conservation. *Plant Genet. Res. Newsl.* 131.
- Jarvis A., Ferguson M., Williams D., Guarino L., Jones P., Stalker H., Valls J., Pittman R., Simpson C. and Bramel P. 2003. Biogeography of wild *Arachis*: Assessing conservation status and setting future priorities. *Crop Sci.* 43: 1100–1108.
- Jones P.G. 1991. The CIAT Climate Database Version 3.41, Machine Readable Dataset. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.
- Jones P.G. and Gladkov A. 1999. FloraMap: A Computer Tool for the Distribution of Plants and Other Organisms in the Wild. CIAT, Cali, Colombia.
- Jones P.G., Beebe S.E., Tohme J. and Galwey N.W. 1997. The use of geographical information systems in biodiversity exploration and conservation. *Biodivers. Conserv.* 6: 947–958.
- Jones P., Guarino L. and Jarvis A. 2002. Computer tools for spatial analysis of plant genetic resources data: 2. FloraMap. *Plant Genet. Res. Newsl.* 130: 1–6.
- Landis J. and Koch G. 1977. The measurements of observer agreement for categorical data. *Biometrics* 33: 159–174.
- Manel S., Williams H. and Ormerod S. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.* 38: 921–931.
- NOAA (National Oceanographic and Atmospheric Administration), 1984. TGPO006 D. Computer Compatible Tape. Boulder, CO, USA.
- Segura S.D., Coppens d'Eeckenbrugge G., Bohorquez A., Ollitrault P. and Tohme J. 2002. An AFLP diversity study of the genus *Passiflora* focusing on subgenus *Tacsonia*. *Genet. Resour. Crop Evol.* 49: 111–123.
- Walker P. and Cocks K. 1991. Habitat: A procedure for modelling a disjoint environmental envelope for a plant or animal species. *Global Ecol. Biogeogr. Lett.* 1: 108–118.